SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis

Year: 2023 | Citations: 3596 | Authors: Dustin Podell, Zion English, Kyle Lacey, A. Blattmann, Tim Dockhorn

---

Abstract

We present SDXL, a latent diffusion model for text-to-image synthesis. Compared to previous versions of Stable Diffusion, SDXL leverages a three times larger UNet backbone: The increase of model parameters is mainly due to more attention blocks and a larger cross-attention context as SDXL uses a second text encoder. We design multiple novel conditioning schemes and train SDXL on multiple aspect ratios. We also introduce a refinement model which is used to improve the visual fidelity of samples generated by SDXL using a post-hoc image-to-image technique. We demonstrate that SDXL shows drastically improved performance compared the previous versions of Stable Diffusion and achieves results competitive with those of black-box state-of-the-art image generators. In the spirit of promoting open research and fostering transparency in large model training and evaluation, we provide access to code and model weights at https://github.com/Stability-AI/generative-models